# Absolute Continuity and Mutual Information for Gaussian Mixtures

C.R. Baker* and A.F. Gualtierotti**

*Department of Statistics
University of North Carolina
Chapel Hill, N.C. 27599-3260

**IDHEAP
University of Lausanne
Lausanne, Switzerland 1015

LISS 28

December, 1988



ABSTRACT

Absolute continuity, process representations, and the Shannon information are considered for problems involving a Gaussian mixture process $(N_t)$, t in $[0,1]$. $N(\omega,t) = A(\omega)G(\omega,t)$ a.e. $dP(\omega)dt$, where $(G_t)$ is a Gaussian process and A is a positive random variable independent of $(G_t)$. Let $(Y_t)$, t in $[0,1]$, be a second process with $\nu_Y$ and $\nu_N$ the measures induced on $\mathbb{R}^{[0,1]}$ and $\mu_Y$ and $\mu_N$ the measures induced on $L_2[0,1]$ (when $(Y_t)$ has paths a.s. in $\mathcal{L}_2[0,1]$). The Cramér-Hida spectral representation and an extension of Girsanov's theorem are used to obtain results on absolute continuity ($\nu_Y \ll \nu_N$ and $\mu_Y \ll \mu_N$) and likelihood ratio in terms of similar results involving a Gaussian mixture local martingale, for which representations are given. These results are then applied to obtain the Shannon mutual information for a communication channel with feedback having $(N_t)$ as additive noise. Capacity is obtained for the no-feedback channel, subject to an average-energy type of constraint.

Research supported by ONR Grant N00014-89-J-1175 and NSF Grant NCR-8713727

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

# REPORT DOCUMENTATION PAGE

| | |
|---|---|
| 1a. REPORT SECURITY CLASSIFICATION<br>UNCLASSIFIED | 1b. RESTRICTIVE MARKINGS |
| 2a. SECURITY CLASSIFICATION AUTHORITY<br><br>2b. DECLASSIFICATION/DOWNGRADING SCHEDULE | 3. DISTRIBUTION/AVAILABILITY OF REPORT<br>Approved for Public Release:<br>Distribution Unlimited |
| 4. PERFORMING ORGANIZATION REPORT NUMBER(S) | 5. MONITORING ORGANIZATION REPORT NUMBER(S) |
| 6a. NAME OF PERFORMING ORGANIZATION<br>Department of Statistics<br>6b. OFFICE SYMBOL (If applicable) | 7a. NAME OF MONITORING ORGANIZATION |
| 6c. ADDRESS (City, State and ZIP Code)<br>University of North Carolina<br>Chapel Hill, North Carolina 27514 | 7b. ADDRESS (City, State and ZIP Code) |
| 8a. NAME OF FUNDING/SPONSORING ORGANIZATION<br>Office of Naval Research<br>8b. OFFICE SYMBOL (If applicable) | 9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER<br>N00014-89-J-1175 |
| 8c. ADDRESS (City, State and ZIP Code)<br>Statistics & Probability Program<br>Arlington, VA 22217 | 10. SOURCE OF FUNDING NOS.<br>PROGRAM ELEMENT NO.: NR<br>PROJECT NO.: 042<br>TASK NO.: 269<br>WORK UNIT NO: SRC 105 |
| 11. TITLE (Include Security Classification)<br>Absolute Continuity and Mutual Information ... | |

12. PERSONAL AUTHOR(S) C.R. Baker and A.F. Gualtierotti

| 13a. TYPE OF REPORT | 13b. TIME COVERED | 14. DATE OF REPORT (Yr., Mo., Day) | 15. PAGE COUNT |
|---|---|---|---|
| TECHNICAL | FROM ________ TO ________ | December 1988 | 28 |

16. SUPPLEMENTARY NOTATION

11. TITLE CONT. for Gaussian Mixtures.

| 17. COSATI CODES | | | 18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number) |
|---|---|---|---|
| FIELD | GROUP | SUB. GR. | Absolute continuity, Shannon information, Gaussian mixtures, Information capacity. |

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

Absolute continuity, process representations, and the Shannon information are considered for problems involving a Gaussian mixture process $(N_t)$, t in [0,1]. $N(\omega,t) = A(\omega)G(\omega,t)$ a.e. $dP(\omega)dt$, where $(G_t)$ is a Gaussian process and A is a positive random variable independent of $(G_t)$. Let $(Y_t)$, t in [0,1], be a second process with $\nu_Y$ and $\nu_N$ the measures induced on $\mathbb{R}^{[0,1]}$ and $\mu_Y$ and $\mu_N$ the measures induced on $L_2[0,1]$ (when $(Y_t)$ has paths a.s. in $L_2[0,1]$). The Cramér-Hida spectral representation and an extension of Girsanov's theorem are used to obtain results on absolute continuity ($\nu_Y \ll \nu_N$ and $\mu_Y \ll \mu_N$) and likelihood ratio in terms of similar results involving a Gaussian mixture local martingale, for which representations are given. These results are then applied to obtain the Shannon mutual information for a communication channel with feedback having $(N_t)$ as additive noise. Capacity is obtained for the no-feedback channel, subject to an average-energy type of constraint.

| | |
|---|---|
| 20. DISTRIBUTION/AVAILABILITY OF ABSTRACT<br>UNCLASSIFIED/UNLIMITED XX SAME AS RPT. XX DTIC USERS ☐ | 21. ABSTRACT SECURITY CLASSIFICATION |
| 22a. NAME OF RESPONSIBLE INDIVIDUAL<br>C.R. Baker | 22b. TELEPHONE NUMBER (Include Area Code)<br>(919) 962-2189<br>22c. OFFICE SYMBOL |

## 0. Introduction

Let I denote the interval [0,1]. $(N_t)$ will denote a real valued stochastic process having index set I and representation $(AG_t)$, where $(G_t)$ is a mean-square continuous Gaussian process with index set I and mean zero, and A is a positive real random variable, independent of $(G_t)$. Let $(Y_t)$ be a second stochastic process with index set I. The measures induced by $(Y_t)$ and $(N_t)$ on $\mathbb{R}^I$ are denoted $\nu_Y$ and $\nu_N$ respectively. The problem first considered here is that of determining conditions for $\nu_Y$ to be absolutely continuous with respect to $\nu_N$ (denoted $\nu_Y << \nu_N$).

The class of processes $(N_t)$ having the above representation is quite large, containing not only Gaussian processes, but also those that are spherically invariant [13] and a large class of $\alpha$-subGaussian processes [7]. In the $\alpha$-subGaussian case, $A^2$ is required to be $\alpha/2$-stable, and in the spherically invariant case $EA^2$ is required to be finite. We define $(N_t)$ to be a Gaussian mixture process and $\nu_N$ a Gaussian mixture measure.

Since G is mean-square continuous, almost all paths of $(N_t)$ belong to $\mathcal{L}_2[0,1]$, and we also consider the absolute continuity problem for measures $\mu_Y$ and $\mu_N$ induced on $L_2[0,1]$. When absolute continuity holds, we obtain the Radon-Nikodym derivative $d\mu_Y/d\mu_N$.

Conditions for absolute continuity and expressions for the Radon-Nikodym derivative when N is Gaussian are given in [4].

The paper has been organized to match the ideas structuring the methods used in obtaining its results. The Cramér-Hida representation [8,11] allows one to consider the noise as the summed output from a set of causal, non-anticipating and non-random linear filters, each of which is driven by a "white noise" process. Assuming that the received signal has paths that are in the



A-1

reproducing kernel Hilbert space (RKHS) of the noise, which means that they are smoother than those of the noise, one can represent the received signal as summed outputs of the same set of linear filters, provided the vector of input signal-plus-noise processes has a law that is absolutely continuous with respect to "white noise" [4]. Thus the problem of the absolute continuity of the inputs (vector processes) is first solved in Section 1 and Section 2, and that for the outputs in Section 3.

The work on absolute continuity w.r.t. a Gaussian mixture has at least two major applications. One is to detection of signals imbedded in noise of this type; see [5], [6], [21] for a discussion of such problems. The second application is to obtain a complete and detailed derivation of the (Shannon) mutual information in an additive feedback channel when the noise is a Gaussian mixture. Even for the Gaussian channel, such a derivation does not exist, although the groundwork has been laid in papers by Kadota, Zakai, and Ziv [15] and by Hitsuda and Ihara [12]. We have previously stated this result without proof [3]. Section 4 contains a detailed derivation. Capacity of the no-feedback channel follows immediately from this expression and a lower bound given in [3]. An information channel is said to be mismatched [2] if the constraint on the transmitted signal is not given directly in terms of the covariance of the channel noise. Such channels constitute the usual case in practical problems. Although the capacity problem has been solved for mismatched no-feedback Gaussian channels [2], the solution is obtained by Hilbert space methods that do not easily carry over to time-continuous feedback channels. The stochastic process formulation given here seems better suited to the feedback channel. Capacity of mismatched channels with feedback comprises one of the major areas of open problems in fundamental information

theory; even for the Gaussian time-discrete channel, a general solution has not been obtained. Thus, it seems particularly important that a complete and general development of the expression for mutual information be available.

All stochastic processes will be defined on a complete probability space $(\Omega, \mathscr{B}, P)$ and will have index set I, unless otherwise stated. A stochastic process whose law is determined by P will be written in simple form, e.g., $(V_t)$. If the law is determined by a probability Q on $(\Omega, \beta)$, then the process will be denoted by $(V_t^Q)$. For a stochastic process $(V_t)$, $\sigma_t^0(V)$ will denote the $\sigma$-field generated by $\{V_s, s \leq t\}$, and $\sigma_t(V)$ its P-completion. $\underline{\sigma}^0(V)$ and $\underline{\sigma}(V)$ are the corresponding filtrations, for example: $\underline{\sigma}^0(V) = \{\sigma_t^0(V), t \in I\}$. $\overline{L_t(V)}$ is the closed linear span in $L_2[P]$ of the set $\{V_s, s \leq t\}$.

For a positive integer $M < \infty$, $\mathscr{R}^M$ will be the Borel $\sigma$-field of $\mathbb{R}^M$ under the product topology. $C_0$ is the set of all the real valued functions defined on I that are continuous and vanish at $t = 0$. $C_0$ is endowed with the sup-norm topology, and $\mathscr{C}_0$ is the resulting $\sigma$-field, also generated by the evaluation maps $\Pi_t$: $\Pi_t(f) = f(t)$, $t \in I$, $f \in C_0$. The Borel $\sigma$-field of $C_0^M$ is $\mathscr{C}_0^M$, the product $\sigma$-field of M copies of $\mathscr{C}_0$. $\mathbb{R}^I$ is the space of real-valued functions defined on I and $\mathscr{R}^I$ the $\sigma$-field in $\mathbb{R}^I$ generated by the cylinder sets, that is, sets of the form

$$\{x: (x(t_1), \ldots, x(t_n)) \in A^n\}, \ n \geq 1, \ t_1, \ldots, t_n \text{ in } I, \ A^n \in \mathscr{R}^n.$$

Let $(\underline{V}_t)$ have paths almost surely in $C_0^M$: it induces a measure $P_{\underline{V}}$ on $\mathscr{C}_0^M$. For a scalar process $(V_t)$, $\upsilon_V$ is the measure induced by $(V_t)$ on $\mathscr{R}^I$, and, if it is measurable, with paths almost surely in $\mathscr{L}_2[0,1]$, the measure induced on $L_2[0,1]$ is denoted $\mu_V$. V will denote the path map: $V[\omega] = \{V(\omega,t), t \in I\}$, or $\underline{V}[\omega] = \{\underline{V}(\omega,t), t \in I\}$.

## 1. Gaussian Mixture Local Martingales

The Gaussian mixture process $(N_t)$ will eventually be defined in terms of a Gaussian mixture local martingale; we discuss processes of this type in this and the following section, and give two useful representations.

Let $\underline{\mathcal{A}} = \{\mathcal{A}_t, t \in I\}$ be a filtration of sub-$\sigma$-fields of $\mathcal{B}$. We make the following assumptions:

A1: Each $\mathcal{A}_t$, $t \in I$, contains the sets of $\mathcal{B}$ having P-measure zero.

A2: $A: \Omega \to \mathbb{R}_+$ is $\mathcal{B}$-measurable and $P\{\omega: A(\omega) > 0\} = 1$.

$\mathcal{M}(M)$ is the family of M×M real matrices. Each matrix in $\mathcal{M}(M)$ is identified with a point in $\mathbb{R}^{M^2}$.

A3: $\Gamma: I \to \mathcal{M}(M)$ has the following properties:

a) $\Gamma(0) = 0$,

b) for $0 \le s < t \le 1$, $\Gamma(t) - \Gamma(s)$ is symmetric and strictly positive definite.

A4: $\underline{B}: \Omega \times I \to \mathbb{R}^M$ is a continuous local martingale for $(\underline{\mathcal{A}}, P)$ and

a) $\underline{B}(\omega,0) = 0$ P-almost surely,

b) $\langle\underline{B}\rangle(\omega,t) = A^2(\omega)\Gamma(t)$, where $\langle\underline{B}\rangle$ is the variation process of $(\underline{B}_t)$.

Thus, for every fixed $\underline{\theta} \in \mathbb{R}^M$, the process $(B_{\theta,t})$ defined by $B_\theta(\omega,t) := \langle\underline{\theta},\underline{B}(\omega,t)\rangle_{\mathbb{R}^M}$ is a continuous local martingale such that $\langle B_\theta\rangle(\omega,t) = A^2(\omega)\langle\underline{\theta},\Gamma(t)\underline{\theta}\rangle_{\mathbb{R}^M}$ [16].

<u>Proposition 1</u>: $\underline{B}$ has the representation $\underline{B}(\omega,t) = A(\omega)\underline{G}(\omega,t)$, $t \in I$, P-almost surely, with $\underline{G}$ a continuous Gaussian local martingale independent of A and such that $\langle\underline{G}\rangle = \Gamma$. A is adapted to $\mathcal{A}_0^+$. $(\underline{B}_t)$ is an $(\underline{\mathcal{A}},P)$ martingale if and only if $EA < \infty$ and an $L_2$-martingale if and only if $EA^2 < \infty$.

$\underline{\text{Remark}}$. In the Introduction, we have begun with a Gaussian process $(G_t)$. In this section, we begin with a vector process $(\underline{B}_t)$ as in A4, and derive a representation in terms of a vector process $(\underline{G}_t)$.

$\underline{\text{Proof}}$: Since $\langle\underline{B}\rangle = A^2\Gamma$ and $\Gamma(t)$ is positive definite for $t > 0$, $A^2(\omega)$ = trace $\langle\underline{B}\rangle(\omega,t)/\text{trace}\ \Gamma(t)$, so that A is $\mathscr{A}_0^+$-measurable.

$(\underline{B}_t)$ is a continuous local martingale, so that $\langle\underline{B}\rangle$ is continuous [16]. Since A is almost surely positive, $\Gamma$ must also be continuous. Define a vector process $(\underline{G}_t)$ such that $\underline{G}(\omega,t)$ is in $\mathbb{R}^M$ by $\underline{G}(\omega,0) = \underline{0}$ and $\underline{G}(\omega,t) = \underline{B}(\omega,t)/A(\omega)$, $t > 0$, $t \in I$. $(\underline{G}_t)$ is then continuous and adapted to $\underline{\mathscr{A}}$. To prove that $(\underline{G}_t)$ is a Gaussian martingale with variation process $\Gamma$, it is sufficient to show that $(\underline{G}_t)$ is a local martingale with variation process $\Gamma$ [16], which is equivalent to establishing that, given any fixed $\underline{\theta}$ in $\mathbb{R}^M$, $G_\theta := \langle\underline{\theta},\underline{G}\rangle_{\mathbb{R}^M}$ is a local martingale with variation process $\langle\underline{\theta},\Gamma\underline{\theta}\rangle_{\mathbb{R}^M}$.

Let now $\{R_n^\theta\}$ be a sequence of stopping times which reduces $B_\theta$, and define the following stopping times:

$$S_n^\theta(\omega) := \inf\{t \geq 0:\ |B_\theta(\omega,t)| > n\},$$

$$T_n^\theta(\omega) := \inf\{t \geq 0:\ |G_\theta(\omega,t)| > n\},$$

$$U_n^\theta := R_n^\theta \wedge S_n^\theta \wedge T_n^\theta.$$

As n increases, each of these stopping times converges almost surely to 1, since $B_\theta$ and $G_\theta$ are each almost surely path continuous on the compact I. Now $G_\theta^{U_n}$ is bounded by n and thus integrable. To see that it is a martingale, there are two cases to consider:

1. $0 < s < t$

It then suffices to write $G_\theta^{U_n^\theta}$ as $B_\theta^{U_n^\theta}/A$ and to take into account that $A$ is adapted to $\mathcal{A}_0^+$, and $B_\theta^{U_n^\theta}$ a martingale.

2 $0 = s < t$

One must then use a limiting argument. Fix $u$ such that $0 < u < t$. It follows as in 1. that

$$E\{G_\theta^{U_n^\theta}(\cdot,t)\,|\,\mathcal{A}_0\} = E\{G_\theta^{U_n^\theta}(\cdot,u)\,|\,\mathcal{A}_0\}.$$

But $G_\theta^{U_n^\theta}$ is almost surely path continuous and thus, almost surely,

$\lim_{u\downarrow 0} G_\theta^{U_n^\theta}(\cdot,u) = 0$. Since furthermore $G_\theta^{U_n^\theta}$ is uniformly bounded,

$\lim_{u\downarrow 0} E\{G_\theta^{U_n^\theta}(\cdot,u)\,|\,\mathcal{A}_0\} = 0$, almost surely.

To verify that $\langle G_\theta\rangle = \langle \underline{\theta}, \Gamma\underline{\theta}\rangle_{\mathbb{R}^M}$, one notes that

$$\{G_\theta^{U_n^\theta}(\cdot,t)\}^2 - \langle \underline{\theta}, \Gamma(t\wedge U_n^\theta)\underline{\theta}\rangle_{\mathbb{R}^M} = \{[B_\theta^{U_n^\theta}(\cdot,t)]^2 - \langle B_\theta\rangle^{U_n^\theta}(\cdot,t)\}/A^2.$$

Since $U_n^\theta \le R_n^\theta$, $B_\theta^{U_n^\theta}$ is a uniformly integrable martingale for fixed n. Since A is adapted to $\mathcal{A}_0^+$, the right side of the last equality is a martingale.

To check that A is independent of $\underline{G}$, fix $0 < s < t_1 < \dots < t_n$, and let

$$I(\alpha,\underline{\alpha}_1,\dots,\underline{\alpha}_n) := E\exp[i\{\alpha A + \sum_{i=1}^{n} \langle \underline{\alpha}_j, \underline{G}(\cdot,t_j)\rangle_{\mathbb{R}^M}\}].$$

Conditioning on $\mathcal{A}_0^+$ and using the fact that $(\underline{G}_t)$ is a Gaussian martingale, one has:

$$I(\alpha,\underline{\alpha}_1,\dots,\underline{\alpha}_n) := E\exp\{i\alpha A\}\, E\exp(i\sum_{i=1}^{M} \langle \underline{\alpha}_j, \underline{G}(\cdot,t_j)\rangle_{\mathbb{R}^M}\}. \qquad \square$$

Proposition 1 is the equivalent, for Gaussian mixture local martingales, of the martingale characterization of Brownian motion. So, essentially the same proof which is valid in the Brownian case [9] applies to Gaussian mixture local martingales and one thus has:

Proposition 2: Let $\mathcal{A}_0 := \sigma_0^+(\underline{B})$ and $\mathcal{A}_t := \sigma_t(\underline{B})$, $t > 0$ and assume that $\mathcal{A}_1 = \mathcal{B}$, $\Gamma(t) = \mathrm{diag}[\Gamma_1(t),\dots,\Gamma_M(t)]$, $t \in I$. Then every local martingale $(V_t)$ for $(\underline{\mathcal{A}},P)$, with paths almost surely in $D[0,1]$, has the representation

$$V(\omega,t) = V_0(\omega) + \sum_{i=1}^{M} \int_0^t V_i(\omega,y) B_i(\omega,dy),$$

with $V_0$ adapted to $\mathcal{A}_0$, each $(V_{i,t})$ $(\underline{\mathcal{A}},P)$-predictable and $\int_0^t V_i^2(\omega,y)\Gamma_i(dy) < \infty$ almost surely for fixed t in I.

## 2. Absolute Continuity With Respect to a Gaussian Mixture Local Martingale

$(\underline{B}_t)$ is the Gaussian mixture local martingale defined in Section 1. In this section, we give sufficient conditions for $P_{\underline{X}} << P_{\underline{B}}$ and $P_{\underline{X}} \sim P_{\underline{B}}$ where $\underline{X}$ is a second M-vector process, and also give an expression for the Radon-Nikodym derivative $dP_{\underline{X}}/dP_{\underline{B}}$. We make the following additional assumptions.

A5: $(\underline{X}_t)$ is an M-component vector process, separable with respect to closed sets.

A6: $(\underline{s}_t)$ is an M-component vector process, $\underline{\mathcal{A}}$-optional and such that $\int_0^1 \langle \underline{s}(\omega,y), \Gamma(dy)\underline{s}(\omega,y)\rangle_{\mathbb{R}^M} < \infty$, almost surely.

A6*: Let $\underline{s}^*: C_0^M \times I \to \mathbb{R}^M$ be $\underline{\mathcal{C}}_0^M$-optional ($\mathcal{C}_{0t}^M$ := $\sigma$-algebra generated by the evaluation maps $\{\Pi(\underline{c},s) = \underline{c}(s),\ s \le t\}$), and such that $\int_0^1 \langle \underline{s}^*(X[\omega],y),\ \Gamma(dy)\underline{s}^*(X[\omega],y)\rangle_{\mathbb{R}^M} < \infty$, almost surely.

A7: For each t in I, $\underline{X}(\omega,t) = \int_0^t \langle \underline{B}\rangle(\omega,dy)\underline{s}(\omega,y) + \underline{B}(\omega,t)$, almost surely.

A7*: For each t in I, $\underline{X}(\omega,t) = \int_0^t \langle\underline{B}\rangle(\omega,dy)\underline{s}^*(X[\omega],y) + \underline{B}(\omega,t)$, almost surely.

A8: Let $\Delta(\omega,t) := \exp\{-\int_0^t \langle\underline{s}(\omega,y),\underline{B}(\omega,dy)\rangle_{\mathbb{R}^M}$

$- \frac{1}{2}\int_0^t \langle\underline{s}(\omega,y),\langle\underline{B}\rangle(\omega,dy)\underline{s}(\omega,y)\rangle_{\mathbb{R}^M}\}$, $D(\omega) := \Delta(\omega,1)$, and assume $ED = 1$.

A8*: Let $\Delta^*$ be obtained by choosing $\underline{s}^*(X[\omega],t)$ for $\underline{s}(\omega,t)$ in A8, $D^*$ be defined accordingly, and assume that $ED^* = 1$.

Proposition 3: Let $dQ \equiv DdP$, $Q_A \equiv Q \circ A^{-1}$, $P_A \equiv P \circ A^{-1}$, $Q_{\underline{X}} \equiv Q \circ \underline{X}^{-1}$, $P_{\underline{B}} \equiv P \circ \underline{B}^{-1}$. Then, with assumptions A1-A8,

a) for any $\mathscr{A}_0^+$-measurable random variable U which is almost surely bounded with respect to P, $E_Q U = E_P U$.

b) $Q_A = P_A$ and $Q_{\underline{X}} = P_{\underline{B}}$.

Proof: A5 and A7 imply almost sure path equality in A7, so that the paths of $(\underline{X}_t)$ are almost surely continuous. By Girsanov's theorem [16], $(\underline{X}_t)$ is a continuous $(\mathscr{A},Q)$-local martingale such that $\langle\underline{X}\rangle^Q = A^2\Gamma$. Consequently, by Proposition 1, $(\underline{X}_t) = (A\underline{G}_t^Q)$, where $(\underline{G}_t^Q)$ is a continuous $(\mathscr{A},Q)$-Gaussian martingale with $\langle\underline{G}^Q\rangle = \Gamma$, and A is independent of $(\underline{G}_t^Q)$ with respect to Q.

Let U be any $\mathscr{A}_0^+$-measurable random variable that is almost surely bounded with respect to P. Fix $t > 0$. Then, because of A8, $\Delta$ is a martingale and

$$E_Q U = E_P\{U\Delta(\cdot,t)\}.$$

But $\Delta$ has also continuous paths and $\lim_{t\downarrow 0}\Delta(\omega,t) = 1$, almost surely with respect to P. Furthermore, still with respect to P, $\Delta(\omega,t) \geq 0$ almost surely, and $E\Delta(\cdot,t) = 1$, for all t in I. Consequently, $\Delta(\cdot,t) \to 1$ in $\sigma(L_1(P),L_\infty(P))$ as $t \downarrow 0$ [9]. Thus:

$$E_Q U = \lim_{t\downarrow 0} E_P\{U\Delta(\cdot,t)\} = E_P U.$$

The equalities $Q_A = P_A$ and $Q_{\underline{X}} = P_{\underline{B}}$ follow directly from this result, using characteristic functions: choose first $e^{i\theta A}$ for U, then clearly, $P_A = Q_A$. Now consider $E_Q V$ (= $E_P V$), where $V = \exp\{i \sum_{j=1}^{n} \langle \underline{\alpha}_j, \underline{X}(\cdot, t_j)\rangle_{\mathbb{R}^M}\}$. Then $P_A = Q_A$ yields:

$$\begin{aligned} E_Q \exp\{i \sum_{j=1}^{n} \langle \underline{\alpha}_j, \underline{X}(\cdot, t_j)\rangle_{\mathbb{R}^M}\} &= \int_0^{\infty} Q_A(da) E_Q \exp\{ia \sum_{j=1}^{n} \langle \underline{\alpha}_j, \underline{G}^Q(\cdot, t_j)\rangle_{\mathbb{R}^M}\} \\ &= \int_0^{\infty} Q_A(da) \exp\{-\frac{a^2}{2} \sum_{j=1}^{n} \langle \underline{\alpha}_j, \Gamma(t_j)\underline{\alpha}_j\rangle_{\mathbb{R}^M}\} \\ &= \int_0^{\infty} P_A(da) E_P \exp\{ia \sum_{j=1}^{n} \langle \underline{\alpha}_j, \underline{G}(\cdot, t_j)\rangle_{\mathbb{R}^M}\} \\ &= E_P \exp\{i \sum_{j=1}^{n} \langle \underline{\alpha}_j, \underline{B}(\cdot, t_j)\rangle_{\mathbb{R}^M}\}. \end{aligned}$$

□

<u>Proposition 4</u>:

Let $\gamma(\underline{c}, y) := \text{trace}\{\underline{\Pi}(\underline{c}, y)\underline{\Pi}(\underline{c}, y)^t - 2\int_0^y \underline{\Pi}(\underline{c}, z)\underline{\Pi}(\underline{c}, dz)^t\}/\text{trace}\,\Gamma(y)$, $\underline{c} \in C_0^M$ and $y \in I$ ($\underline{\Pi}(\underline{c}, y) := \underline{c}(y)$). Then:

a) A1-A7 imply $P_{\underline{X}} \ll P_{\underline{B}}$ and A1-A8 imply $P_{\underline{X}} \sim P_{\underline{B}}$;

b) A1-A5 and A6*-A8* imply $P_{\underline{X}} \sim P_{\underline{B}}$ and, with respect to $P_{\underline{B}}$, for almost every $\underline{c}$ in $C_0^M$,

$$\frac{dP_{\underline{X}}}{dP_{\underline{B}}}(\underline{c}) = \exp\{\int_0^1 \langle \underline{s}^*(\underline{c}, y), \underline{\Pi}(\underline{c}, dy)\rangle_{\mathbb{R}^M} - \frac{1}{2}\int_0^1 \langle \underline{s}^*(\underline{c}, y), \langle\underline{\Pi}\rangle(\underline{c}, dy)\underline{s}^*(\underline{c}, y)\rangle_{\mathbb{R}^M}\},$$

where $\langle\underline{\Pi}\rangle(\underline{c}, y) = \gamma(\underline{c}, y)\Gamma(y)$ [20].

<u>Proof</u>: The result follows from the method used [19] when $(\underline{B}_t)$ is the Wiener process, together with Proposition 3.

Remark. Sufficient conditions for A8 and A8* are respectively [19]:

$$E \exp\{\tfrac{1}{2}\int_0^1 \langle \underline{s}(\omega,y), \langle \underline{B}\rangle(\omega,dy)\underline{s}(\omega,y)\rangle_{\mathbb{R}^M}\} < \infty,$$

$$P_{\underline{B}}\{\int_0^1 \langle \underline{s}^*(\underline{c},y), \langle \underline{\Pi}\rangle(\underline{c},dy)\underline{s}^*(\underline{c},y)\rangle_{\mathbb{R}^M} < \infty\} = 1.$$

Assuming that $E \int_0^1 \langle \underline{s}(\cdot,y), \Gamma(dy)\underline{s}(\cdot,y)\rangle_{\mathbb{R}^M} < \infty$, one can obtain a representation of the form A7* starting from a representation of the form A7. These results are proved as in the Wiener process case [19].

## 3. Absolute Continuity With Respect to a Gaussian Mixture Process

In this section we return to the problem discussed in the Introduction, treating a Gaussian mixture process $(AG_t)$, where $(G_t)$ is a zero-mean mean-square-continuous Gaussian process. G has a proper canonical representation [11], which we assume to have the form

$$G(\omega,t) = \int_0^t \langle \underline{F}(t,x), \underline{G}(\omega,dx)\rangle_{\mathbb{R}^M} . \tag{1}$$

where $\underline{F}$ and $(\underline{G}_t)$ have respective components $F_i$ and $(G_{i,t})$, $1 \le i \le M$.

Without the assumption that $(G_t)$ is Gaussian, the processes $(G_{i,t})$ are zero-mean orthogonal-increment processes, mutually orthogonal and mean-square continuous. Their non-decreasing variances $EG_i^2(\cdot,t)$ define Borel measures $\Gamma_i$ on I in the usual way: $\Gamma_i(a,b] = EG_i^2(\cdot,b) - EG_i^2(\cdot,a)$; moreover $\Gamma_{i+1} \ll \Gamma_i$, $1 \le i \le M-1$. Each $F_i$: $I \times I \to \mathbb{R}$ is Borel-measurable, $F_i(t,x) = 0$ for $x > t$, and $\sum_{i=1}^{M} \int_0^1\int_0^1 F_i^2(t,s)\Gamma_i(ds)dt < \infty$ $M \le \infty$ is the multiplicity of G.

The proper canonical representation (1) has the property that $\overline{L}_t(G) = \overline{L}_t(\underline{G})$, for t in I, and that $\sum_{i=1}^{M} \int_0^1 F_i(t,s)g_i(s)\Gamma_i(ds) = 0$, all t in I,

if and only if $g_i = 0$ in $L_2[\Gamma_i]$, $1 \leq i \leq M$. Thus $\{F_i(t,\cdot),\ t \in I\}$ spans $L_2[\Gamma_i]$.

The representation (1) is an equality in the mean-square sense, thus holds P-almost everywhere for each fixed t in I. However, taking both sides of (1) separable with respect to closed sets means that we can assume path equality P-almost everywhere.

The assumption that $(G_t)$ is Gaussian further implies that $\{(G_{i,t}),\ i \leq M\}$ are a Gaussian family; thus the the component processes are mutually independent, have independent increments, and hence can be assumed to be path continuous. Moreover equality of $\overline{L}_t(G)$ and $\overline{L}_t(\underline{G})$ implies that of $\sigma_t(G)$ and $\sigma_t(\underline{G})$ for all t in I. Each $(G_{i,t})$, $1 \leq i \leq M$, is a martingale with respect to $\underline{\sigma}(G)$ and with respect to $\underline{\sigma}(G) \vee \underline{\sigma}(V)$, where V is any stochastic process independent of G.

M is the *multiplicity of* $(G_t)$, and it is assumed throughout that $M < \infty$. This restriction is due to a similar restriction in [4], on which part of these results depend.

We assume that support$[P_{\underline{G}}] = C_0^M$ and that support$[\mu_G] = L_2[0,1]$. Since these measures are Gaussian, their supports are closed linear manifolds equal to the closure of the ranges of their covariance operators [14]. One can thus always work with this subspace; it preserves the original linear space structure under the original norm (and inner product, for $L_2[0,1]$). As a consequence of these assumptions, one can take (in the proofs) each $\Gamma_i$ to give positive measure to each interval (a,b), b > a.

Let V be a mean square continuous process: its reproducing kernel Hilbert space (RKHS) is denoted $H_V$, and its covariance by $C_V$. All elements in $H_V$ are continuous functions defined on I. $C_V$ is the kernel of an integral operator $R_V$

on $L_2[0,1]$ which is trace-class: the covariance operator of V, $R_V$, defines a Hilbert space of $L_2[0,1]$ elements, denoted $K_V$, consisting of range($R_V^{\frac{1}{2}}$) together with the inner product

$$(k_1,k_2)_V = \Sigma_n \langle k_1,u_n\rangle\langle k_2,u_n\rangle/\zeta_n,$$

where $(\zeta_n)$ and $\{u_n, n\geq 1\}$ are the non-zero eigenvalues and associated orthonormal eigenvectors of $R_V$ and $\langle\cdot,\cdot\rangle$ is the $L_2[0,1]$ inner product. For $L_2[\Gamma_i]$, $1 \leq i \leq M$, already defined, H will denote $\bigoplus_{i=1}^{M} L_2[\Gamma_i]$. Since it is assumed that support$[\mu_G] = L_2[0,1]$, the eigenvalues $(\lambda_n)$ of $R_G$ are all non-zero and the associated orthonormal eigenvectors are complete in $L_2[0,1]$.

Let A be a positive real-valued random variable independent of the process G, and define $(N_t) = (AG_t)$. $(N_t)$ has sample paths almost surely in $\mathcal{L}_2[0,1]$, and its path-map N defines, when taking equivalence classes, a $\mathfrak{F}/\mathbb{B}(L_2[0,1])$ measurable map, where $\mathbb{B}(L_2[0,1])$ is the Borel $\sigma$-field of $L_2[0,1]$. Although $EN_t^2$ is not assumed to be finite, so that $(N_t)$ may not have a covariance function, one can nevertheless consider $H_{N/A}$, the conditional RKHS of $(N_t)$ given A. Note that $H_{N/A} = H_G$ a.s., since $A > 0$ a.s. We assume that $\beta$ is the smallest $\sigma$-field containing $\underline{\sigma}(N)$: $\beta \equiv \sigma_1(N)$.

$(\underline{B}_t)$ is the Gaussian mixture local martingale defined by $(\underline{B}_t) = (A\underline{G}_t)$. $(N_t)$ has the representation:

$$N(\omega,t) = \int_0^t \langle \underline{F}(t,y),\underline{B}(\omega,dy)\rangle_{\mathbb{R}^M}.$$

The fact that this is derived from the proper canonical representation for $(G_t)$ implies the following for F:

Let $(P_t f)(s) = I_{[0,t]}(s)f(s)$, $L_2^t[\Gamma_i] = P_t L_2[\Gamma_i]$, $L_2^t[\Gamma] = \bigoplus_{i=1}^{M} L_2^t[\Gamma_i]$.

Then:

$\alpha$. $\{F_i(s,\cdot),\ s \le t\}$ spans $L_2^t[\Gamma_i]$.

$\beta$. $\{\underline{F}(s,\cdot),\ s \le t\}$ spans $L_2^t[\Gamma]$.

Since each $(G_{i,t})$ is Gaussian and has orthogonal increments, it is a continuous $L_2$-martingale WRT to the $\sigma$-algebras generated by $(N_t)$ and $(\underline{B}_t)$ and the sets of measure zero.

It is assumed that $(S_t)$ has paths which are a.s. in the RKHS of $(G_t)$ and that it is adapted to the filtration $\sigma(\underline{N})$. Then, if $Y_t = S_t + N_t$, a.s. for all fixed t, Y has the representation [4]

$$Y(\omega,t) = \int_0^t \langle \underline{F}(t,y), \underline{X}(\omega,dy)\rangle_{\mathbb{R}^M},$$

where $\underline{X}(\omega,t) = \int_0^t \Gamma(dx)\underline{s}(\omega,x) + \underline{B}(\omega,t)$ and $\underline{s}$ is such that $S(\omega,t) = \int_0^t \langle \underline{F}(t,y), \Gamma(dy)\underline{s}(\omega,y)\rangle_{\mathbb{R}^M}$.

Let the eigenvalues and associated c.o.n. eigenvectors of $R_G$ be denoted by $(\lambda_n)$ and $\{e_n,\ n \ge 1\}$. Let $\underline{m}$: $L_2[0,1] \to C_0^M$ be defined by $m_i(h,t) = \Sigma_{n\ge 1}\langle H_{i,t},\ e_n\rangle\langle h,e_n\rangle/\lambda_n$ where $H_{i,t}[y] = \int_0^t F_i(y,u)\Gamma_i(du)$. One then has:

<u>Proposit'on 5</u>: Let $\nu_Y$ and $\nu_N$ be the measures induced on $(\mathbb{R}^I, \mathfrak{R}^I)$ by Y and N respectively, and let $\mu_Y$ and $\mu_N$ be the measures induced on $(L_2[0,1], \mathbb{B}(L_2[0,1]))$ by the same processes. Then

a) A1-A7 imply $\nu_Y << \nu_N$ and $\mu_Y << \mu_N$;

b) Furthermore $\dfrac{d\mu_Y}{d\mu_N}[h] = \dfrac{dP_{\underline{X}}}{dP_{\underline{B}}}\left[\underline{m}(h)\right]$ $\mu_N$-almost every h in $L_2[0,1]$.

<u>Proof</u>:

a) This result is a consequence of Proposition 4 above and Theorem 1 of [4].

b) To obtain the Radon-Nikodym derivative, one writes:

$$P\{\underline{B} \in C_1,\ N \in C_2\} = P\{A\underline{G} \in C_1,\ AG \in C_2\}$$

$$= \int_0^\infty P_A(da)\ P\{a\underline{G} \in C_1,\ aG \in C_2\}$$

$$= \int_0^\infty P_A(da) \int_{C_2} P_{aG}(dh)\ P\{a\underline{G} \in C_1 | aG = h\}.$$

Now, from [4], we have that $P\{a\underline{G} \in C_1 | aG = h\}$ is a point mass concentrated at the function $\underline{m}^a(h,\cdot)$ with components

$$m_i^a(h,t) := \sum_{n=1}^{\infty} \langle H_{i,t}^a, e_n^a\rangle\langle h, e_n^a\rangle/\lambda_n^a,$$

where, if $R_G^a$ is the covariance operator of aG,

$$R_G^a e_n^a = \lambda_n^a e_n^a,$$

and $H_{i,t}^a(x) = \int_0^t F_i(x,y)a^2\Gamma_i(dy) = a^2 H_{i,t}(x)$. Now $R_G^a = a^2 R_G$, since $R_G e_n = \lambda_n e_n$, $e_n^a = e_n$ and $\lambda_n^a = a^2\lambda_n$. Consequently,

$$\underline{m}^a(h,t) = \underline{m}(h,t).$$

Thus for all a

$$P\{a\underline{G} \in C_1 | aG = h\} = I_{C_1}\{\underline{m}[h]\}),$$

so that

$$P\{\underline{B} \in C_1 | N = h\} = I_{C_1}\{\underline{m}[h]\}. \qquad \square$$

An explicit representation of $d\mu_Y/d\mu_N$ based on Proposition 5 requires a representation of $dP_{\underline{X}}/dP_{\underline{B}}$. But rewriting S to be in the RKHS of N, and applying Proposition 4 yields:

$$\frac{dP_X}{dP_{\underline{B}}}(c) = \exp\Big[\int_0^1 \langle (\underline{s}^*(\underline{c},y)/\gamma^2(\underline{c},y)),\ \underline{\Pi}(\underline{c},dy)\rangle_{\mathbb{R}^M}$$

$$- \tfrac{1}{2}\int_0^1 \langle (\underline{s}^*(\underline{c},y)/\gamma(\underline{c},y)),\ \Gamma(dy)(\underline{s}^*(\underline{c},y)/\gamma(\underline{c},y))\rangle_{\mathbb{R}^M}\Big].$$

This requires of course that $(\underline{X}_t)$ have a stochastic differential equation expression, but this can be obtained as indicated in the Remark following Proposition 4.

Remark. Proposition 3 and the preceding discussion show that Theorem 2 of [4] can be adapted to yield a necessary condition for absolute continuity.

## 4. Computation of Mutual Information For the Channel With Feedback and Spherically-Invariant Noise

### 4.1. Channel Model

We now proceed to derive an expression for the mutual information of a communications channel with feedback, where the noise is additive and spherically-invariant. We begin with a precise description of the channel. The noise, channel output, and message probabilities will all be defined on $(\mathbb{R}^I, \mathfrak{R}^I)$, and all will be denoted by capital P's: $P_N$, $P_Y$, $P_J$. The joint measure defined by $(J,Y)$ and its product measure is then defined on $\mathfrak{R}^I \times \mathfrak{R}^I$, and denoted by $P_{J,Y}$ and $P_J \otimes P_Y$.

a. Assumptions on the noise. As in Section 3, we assume that N is m.s. continuous, vanishes a.s. at $t = 0$ and has a Cramér-Hida representation of multiplicity $M < \infty$. We also assume that $EA^2 = 1$.

b. Assumptions on the message. The message is a real-valued stochastic process $(J_t)$, defined on $(\Omega,\beta)$, and independent of $(N_t)$ (w.r.t. P measure). Since $\sigma_t(N) = \sigma_t(\underline{B}) = \sigma(A) \vee \sigma_t(G)$ for $t \in [0,1]$ because of the proper

canonical represention, $(J_t)$ is independent of A and independent of $(G_t)$.

c. Assumptions on the channel. Let $\Phi$ denote the family of maps $\underline{f}$: $[0,1] \to \mathbb{R}^2$, and let $\underline{\mathcal{F}}$ be the filtration of $\Phi$ generated by the evaluation maps. S is the transmitted signal, and Y is the channel output. $\underline{\Psi}$ will denote a two-component vector with elements $\Psi_1 = J$ and $\Psi_2 = Y$, such that $S(\omega,t) = T[\underline{\Psi}(\omega)](t)$ for $\omega$ in $\Omega$, t in $[0,1]$. T: $\Phi \times [0,1] \to \mathbb{R}$. The following assumptions are also made:

1. T is adapted to $\underline{\mathcal{F}}$.
2. For each fixed t in $[0,1]$, $Y(\omega,t) = T[\underline{\Psi}(\omega)](t) + N(\omega,t)$ a.e. $dP(\omega)$.
3. $T[\underline{\Psi}(\omega)] \in RKHS(N)$, a.e. $dP(\omega)$.
4. $E\|T[\underline{\Psi}(\omega)]\|_N^2 < \infty$, where $\|x\|_N$ is the $H_N$ norm of x.
5. Y is adapted to $\underline{\sigma}^o(J) \vee \underline{\sigma}^o(N)$.
6. $P_{Y^v} \sim P_N$ a.e. $dP_J(v)$, where $P_{Y^v}$ is the probability on $(\mathbb{R}^I, \mathfrak{R}^I)$ defined by the process Y when J is fixed and J = v.

4.2. Discussion

Some important aspects of the mathematical and physical problems are reflected in our choice of constraints. First, the Cramér-Hida representation is well-known to be very difficult to determine, in general. Thus, any reasonable and useful set of constraints should not assume knowledge of this representation. Similarly, although the covariance of N can be reasonably assumed to be known, the probability distribution of A is, for practical purposes, difficult or impossible to determine. In applications, a mapping T as above will be used. Thus, we state all assumptions in terms of this T, the obvious measurability properties of Y, and the RKHS of N. We do assume finite Cramér-Hida multiplicity, because much of our work is based on [4].

Assumptions 3 and 4 are made to ensure that the capacity will be finite. That is, in the Gaussian channel without feedback, a necessary and sufficient condition for finite mutual information is that $E\|S[\omega]\|_N^2 < \infty$ [1]. Thus, if such a condition is not imposed on the transmitted signal, then for almost all a, one has that the mutual information of (S, S+aG) is infinite without using feedback, and it follows that the mutual information of (S, S+N) is infinite.

A few remarks may be useful in order to give some perspective to the results obtained here, including a contrast with other versions, in particular those of Kadota, Zakai and Ziv (KZZ) [15] and those of Hitsuda and Ihara (HI) [12].

KZZ assume the noise in the channel is the Wiener process, HI that it is a Gaussian process with a Cramér-Hida representation for which the variances are absolutely continuous with respect to Lebesgue measure. Although this assumption is not explicitly stated by HI, it seems necessary to their development (cf. Prop. 1 of [12]). It would usually be extremely difficult to verify. In the present paper it is assumed that the noise process is spherically invariant, that is, a scale mixture of Gaussian processes. The basic idea of the HI paper is to extend the KZZ result to any Gaussian process having purely continuous Cramér-Hida spectral representation by using the spectral representation in conjunction with Girsanov's theorem. That approach is also used here. A restriction on the spectral multiplicity of the noise has been introduced, for two reasons: there is no stochastic calculus for martingales in $\mathbb{R}^\infty$, and there has been no attempt on our part to extend by other means the formula for mutual information to the case of infinite multiplicity. KZZ assume that the transmitted signal can be given an explicit functional representation on the space obtained as the product of the message and the noise space. HI assume only smoothness and adaptation properties.

A necessary condition for finite mutual information between the message and the channel output is that the joint measure of message and output be absolutely continuous with respect to their product measure. Having a functional representation for the transmitted signal, KZZ can give conditions on the signal that insure this absolute continuity. This is no longer possible when it is only assumed that the transmitted signal belongs to a particular family, the RKHS of the noise. Thus, one must <u>assume</u> a condition ensuring the absolute continuity. Such an assumption is not stated in HI, who instead refer to the Russian version of [18]. However, in the English version of [18], we are unable to verify that $P_{J,Y} << P_J \otimes P_Y$ using only assumption 3 above (the assumption used by HI), or even with the stronger assumption 4.

Thus, we make assumption 6. It will be seen (Prop. 6 below) that this is enough to give $P_{J,Y} << P_J \otimes P_Y$. Sufficient conditions for $P_{Y^v} \sim P_N$ a.e. $dP_J(v)$ are given in Section 3 above.

In addition to extending the HI results to all m.s. continuous Gaussian noise processes having finite Cramér-Hida multiplicity, and to spherically-invariant processes with finite multiplicity, we also use a substantially different approach. This is in part based on the results of [4], which enable us to bypass some steps in the HI development that are not clear to us.

As in Section 3, there exists a vector process $(\underline{X}_t)$ having paths a.s. in $C_0^M$ and such that $Y(\omega,t) = \int_0^t \langle \underline{F}(t,x), X(\omega,dx)\rangle_{\mathbb{R}^M}$, with $\underline{s}$ being predictable, square-integrable, and adapted to $\underline{\sigma}(\underline{\Psi})$. The basic approach (as in HI) is to compute the mutual information $I(J,Y)$ by showing that it is equal to $I(J,X)$ and to then work with the latter. It is thus essential that $P_{J,\underline{X}} << P_J \otimes P_{\underline{X}}$, and we shall also need the fact that $\underline{\sigma}(Y) = \underline{\sigma}(X)$.

Proposition 6. Under the preceding assumptions, $P_{J,Y} \sim P_J \otimes P_Y$, $P_{J,\underline{X}} \sim P_J \otimes P_{\underline{X}}$, and $\underline{\sigma}(Y) = \underline{\sigma}(\underline{X})$.

Remark. The proof of equivalence of the measures does not require assumption c.4 of Section 4.1.

Proof. Since Y is adapted to $\underline{\sigma}^o(J) \vee \underline{\sigma}^o(N)$, there exists [10] a function $g: \mathbb{R}^I \times \mathbb{R}^I \to \mathbb{R}^I$, measurable, such that for each fixed t in [0,1], $Y(\omega,t) = \pi_t g[J(\omega),N(\omega)]$ a.e. $dP(\omega)$. If $Y^v$ denotes Y when J is fixed and $J = v$, then $Y^v(\omega,t) = \pi_t g[v,N(\omega)]$ a.e. dP, each fixed t. Thus $P_Y = [P_J \otimes P_N] \circ g^{-1}$ and $P_{Y^v} = P_N \circ g_v^{-1}$, $g_v: y \to g(v,y)$. From Assumption c.6 and Theorem 1 of [1], this gives $P_{J,Y} \sim P_J \otimes P_Y$. From Theorem 7 of [4], and proceeding as in the proof of Prop. 5 above, there exists $\underline{m}_0: \mathbb{R}^T \to C_0^M$, measurable, such that $\frac{dP_Y}{dP_N}(y) = \frac{dP_{\underline{X}}}{dP_{\underline{B}}}(\underline{m}_0 y)$ a.e. $dP_N(y)$ and (Theorem 6 of [4]) $P_{\underline{B}} = P_N \circ \underline{m}_0^{-1}$. Thus, for A in $\mathcal{C}^M$,

$$P_{\underline{X}}(A) = \int_A \frac{dP_{\underline{X}}}{dP_{\underline{B}}}(y)dP_{\underline{B}}(y) = \int_{\underline{m}_0^{-1}(A)} \frac{dP_{\underline{X}}}{dP_{\underline{B}}}(\underline{m}_0 y)dP_N(y)$$

$$= \int_{\underline{m}_0^{-1}(A)} \frac{dP_Y}{dP_N}(y)dP_N(y) = P_Y \circ \underline{m}_0^{-1}(A).$$

Thus, $P_{J,Y} \sim P_J \otimes P_Y \Rightarrow P_{J,\underline{X}} \sim P_J \otimes P_{\underline{X}}$.

To show that $\underline{\sigma}(Y) = \underline{\sigma}(\underline{X})$, we again note that (as in Section 3) $Y(\omega,t) = \int_0^t \langle \underline{F}(t,x), \underline{X}(\omega,dx)\rangle_{\mathbb{R}^M}$ where $\underline{X}(\omega,t) = \int_0^t \Gamma(dx)\underline{s}(\omega,x) + \underline{B}(\omega,t)$ with $\underline{s}$ predictable, square-integrable and adapted to $\underline{\sigma}(\underline{\Psi})$. This representation of Y yields $\underline{\sigma}(Y) \subset \underline{\sigma}(\underline{X})$.

To prove that $\underline{\sigma}(\underline{X}) \subset \underline{\sigma}(Y)$, it is sufficient to show that the closed linear span of $\{\underline{X}_s, s \leq t\}$ is contained in $\overline{\text{span}\{Y_s, s \leq t\}}$ for all $t \in [0,1]$. For this, let $I_t$ be the indicator of [0,t], let $\underline{\theta}$ be any element of $\mathbb{R}^M$, and set

$\underline{\xi}_n(x) = I_t(x)\underline{\theta} - \sum_{i=1}^{n} \alpha_i \underline{F}(s_i, x)$, where $s_1 \leq s_2 \leq \cdots \leq s_n \leq t$ and $\{\alpha_i,\ i \leq n\}$ is one set of n real scalars. Then

$$\langle \underline{\theta},\ \underline{X}(\omega,t)\rangle_{\mathbb{R}^M} - \sum_{i=1}^{n} \alpha_i Y(\omega, s_i) = \int_0^1 \langle [I_t(x)\underline{\theta} - \sum_{i=1}^{n} \alpha_i \underline{F}(s_i, x)],\ \underline{X}(\omega, dx)\rangle_{\mathbb{R}^M}$$

$$= \int_0^1 \langle \underline{\xi}_n(x),\ \Gamma(dx)\underline{s}(\omega,x)\rangle_{\mathbb{R}^M} + \int_0^1 \langle \underline{\xi}_n(x),\ \underline{B}(\omega,dx)\rangle_{\mathbb{R}^M}.$$

Thus,

$$E\{\langle \underline{\theta}, \underline{X}(\cdot,t)\rangle_{\mathbb{R}^M} - \sum_{i=1}^{n} \alpha_i Y(\cdot, s_i)\}^2$$

$$\leq 2\left\{E\left[\int_0^1 \langle \underline{\xi}_n(x),\ \Gamma(dx)\underline{s}(\cdot,x)\rangle_{\mathbb{R}^M}\right]^2 + E\left[\int_0^1 \langle \underline{\xi}_n(x),\ \underline{B}(\cdot,dx)\rangle_{\mathbb{R}^M}\right]^2\right\}.$$

Now, using $\|\cdot\|_N$ to denote the norm in RKHS(N),

$$E\left[\int_0^1 \langle \underline{\xi}_n(x),\ \Gamma(dx)\underline{s}(\cdot,x)\rangle_{\mathbb{R}^M}\right]^2$$

$$\leq E\left[\int_0^1 \langle \underline{\xi}_n(x),\ \Gamma(dx)\underline{\xi}_n(x)\rangle_{\mathbb{R}^M} \int_0^1 \langle \underline{s}(\cdot,x),\ \Gamma(dx)\underline{s}(\cdot,x)\rangle_{\mathbb{R}^M}\right]$$

$$= E\|\underline{\xi}_n\|^2_{L_2(\Gamma)} \|\underline{T}\circ\underline{\Psi}[\cdot]\|^2_N = \|\underline{\xi}_n\|^2_{L_2[\Gamma]} E\|\underline{T}\circ\underline{\Psi}[\cdot]\|^2_N.$$

Since $E[\int_0^1 \langle \underline{\xi}_n(x),\ \underline{B}(\cdot,dx)\rangle_{\mathbb{R}^M}]^2 = \|\underline{\xi}_n\|^2_{L_2[\Gamma]}$, it follows that

$$E\{\langle \underline{\theta}, \underline{X}(\cdot,t)\rangle_{\mathbb{R}^M} - \sum_{i=1}^{n} \alpha_i Y(\cdot, s_i)\}^2 \leq \|\underline{\xi}_n\|^2_{L_2[\Gamma]}\{1 + E\|T\circ\underline{\Psi}[\cdot]\|^2_N\}.$$

Since $\underline{F}$ arises in a proper canonical representation, $(\alpha_1, \ldots, \alpha_n)$ and $(s_1, \ldots, s_n)$ (where $s_i \leq t,\ 1 \leq i \leq n$) can be chosen so that

$$\lim_n \|\underline{\xi}_n\|^2_{L_2[\Gamma]} = 0.$$

This completes the proof that $\underline{\sigma}(Y) = \underline{\sigma}(X)$. □

We now have that $I(J,Y) = I(J,\underline{X})$. Thus, the computation of $I(J,Y)$ is reduced to that of $I(J,X)$ and the latter requires an explicit expression for

$$D_{J,\underline{X}} \equiv \frac{dP_{J,\underline{X}}}{dP_J \otimes dP_{\underline{X}}} .$$

4.3. <u>Computation of $D_{J,\underline{X}}$</u>

In this section we adopt the procedure used by KZZ, and must therefore "lift" the problem as defined on the space $\Omega$ to the product of the path space of message and noise. Consequently, this computation is based on the following formula, which takes into account that J and $\underline{B}$ are independent (assumption 4.1.b).

$$D_{J,\underline{X}}(J,\underline{X}) = \frac{(dP_{J,\underline{X}}/dP_{J,\underline{B}})(J,\underline{X})}{(dP_J \otimes dP_{\underline{X}}/dP_{J,\underline{B}})(J,\underline{X})} = \frac{(dP_{J,\underline{X}}/dP_{J,\underline{B}})(J,\underline{X})}{(dP_{\underline{X}}/dP_{\underline{B}})(\underline{X})} \equiv \frac{D_{J,\underline{X},\underline{B}}(J,\underline{X})}{D_{\underline{X},\underline{B}}(\underline{X})} .$$

(a) <u>Representation of</u> $(\underline{X}_t)$. $D_{J,\underline{X},\underline{B}}$ is obtained by showing that it is equal to $(dP_{\underline{X}|J}/dP_{\underline{B}}) \equiv D^c_{J,\underline{X},\underline{B}}$, where $P_{\underline{X}|J}$ is the conditional law for $\underline{X}$ given J, and then by computing $D^c_{J,\underline{X},\underline{B}}$. This is done by "lifting" the equation for $\underline{X}$ onto product space in order to obtain a "decoupling" of the effect of J and N. To do this, one must give an explicit functional representation for $(\underline{X}_t)$.

Let $\Lambda$ be the family of maps $g: [0,1] \to \mathbb{R}^{M+1}$, where $g_0 \in \mathbb{R}^{[0,1]}$ and $g_i \in C_0[0,1]$, $1 \leq i \leq M$. $\underline{\mathcal{G}}$ is the filtration of $\Lambda$ generated by the evaluation maps. $(\underline{x}_t)$ will denote the vector stochastic process with components $x_0 \equiv J$, and $(x_{i,t}) \equiv (X_{i,t})$ for $1 \leq i \leq M$. $(\underline{X}_t)$ then has the representation, a.e. $dP(\omega)$,

$$\underline{X}(\omega,t) = \int_0^t \Gamma(dx)\underline{\tau}(\underline{x}[\omega],x) + \underline{B}(\omega,t)$$

for each fixed t. $\underline{\tau}$ is $\underline{\mathcal{G}}$-predictable [19]. In fact, we know that

i) $(\underline{s}_t)$ is $\underline{\sigma}(\underline{\Psi})$-predictable.

ii) $\underline{\sigma}(\underline{\Psi}) = \sigma(J) \vee \underline{\sigma}(Y)$.

iii) $\underline{\sigma}(Y) = \underline{\sigma}(\underline{X})$.

Thus, $(\underline{s}_t)$ is $\underline{\sigma}(\underline{x})$ predictable, so that $(\underline{s}_t)$ cannot be distinguished (pathwise) from a process which is $\underline{\sigma}^{o}(\underline{x})$-predictable; $\underline{s}$ can be written as $\underline{s}(\omega,t) = \underline{\tau}(\underline{x}[\omega],t)$ ([4, Lemma 2] and [19, p. 67]).

(b) <u>Representation of $\underline{X}$ on function space</u>. Let $q_1$ and $q_2$ be the projections onto the first and second components of $\Phi$. On $\mathcal{F}_1$ (defined in 4.1.c), consider the probability $P_J \otimes P_N$, induced by $\underline{\theta}$: $\Omega \times [0,1] \to \mathbb{R}^2$ with components $\theta_1 \equiv J$ and $\theta_2 \equiv N$. Let $\overline{\mathcal{F}}_1$ be the completion of $\mathcal{F}_1$ with respect to $P_J \otimes P_N$, with $\overline{P_J \otimes P_N}$ the extension of $P_J \otimes P_N$ to $\overline{\mathcal{F}}_1$. Finally, let $\overline{\mathcal{F}}_t$ be the filtration generated by $\mathcal{F}_t$ and the sets of $\overline{\mathcal{F}}_1$ whose $\overline{P_J \otimes P_N}$ measure is zero. $\overline{\mathcal{F}}$ will then denote the resulting filtration of $\Phi$.

With these definitions, there exist processes $\underline{X}_\Phi$, $\underline{B}_\Phi$, and $A_\Phi$ defined on the probability space $(\Phi, \overline{\mathcal{F}}_1, \overline{P_J \otimes P_N})$ which are adapted to $\overline{\mathcal{F}}$ and which have the following properties.

(1) $(\underline{B}_{\Phi,t})$ is a continuous local martingale such that (a.e. $dP(\omega)$, all $t$ in $[0,1]$):

- $\underline{B}_\Phi(\underline{\theta}[\omega],t) = \underline{B}(\omega,t)$
- $(\underline{B}_{\Phi,t})$ depends on $(N_t)$ only (not on $(J_t)$)
- $\langle \underline{B}_\Phi \rangle(\underline{f},t) = A^2_\Phi(\underline{f})\Gamma(t)$
- $A_\Phi(\underline{\theta}[\omega]) = A(\omega)$
- $A_\Phi$ depends on $(N_t)$ only (not on $(J_t)$).

(2) $\underline{X}_\Phi(\theta[\omega],t) = \underline{X}(\omega,t)$.

(3) $\underline{X}_\Phi(\underline{f},t) = \int_0^t \Gamma(dx)\underline{\tau}_\Phi\left[\begin{bmatrix} q_1 \\ \underline{X}_\Phi \end{bmatrix}(\underline{f},x)\right] + \underline{B}_\Phi(\underline{f},t)$

where $\underline{\tau}_\Phi$ is predictable for $\overline{\mathcal{F}}$ and $\underline{\tau}_\Phi\left[\begin{bmatrix} q_1 \\ \underline{X}_\Phi \end{bmatrix} \circ \theta[\omega],t\right] = \underline{\tau}(\underline{x}[\omega],t) = \underline{s}(\omega,t)$.

The validity of these statements is seen as follows. $(\underline{X}_t)$ is continuous and adapted to $\underline{\sigma}(Y)$, so that it is adapted to $\underline{\sigma}(J) \vee \underline{\sigma}(N)$. $(\underline{X}_t)$ is thus predictable, so that it cannot be distinguished (pathwise) from a process that is predictable for $\underline{\sigma}^{\circ}(J) \vee \underline{\sigma}^{\circ}(N)$ [17]. It then has, as in (a), a functional representation (lemma 2 of [4]) as given in (2) above. The same argument applies to $(\underline{B}_t)$ and $\underline{\tau}$. Moreover, since $A(\omega) = \tau(\underline{B}[\omega],t)$, $t > 0$, $A_{\Phi}$ is given by $A_{\Phi}(\underline{f}) = \tau(\underline{B}_{\Phi}[\underline{f}],t)$, $t > 0$.

(c) Computation of $D_{J,\underline{X},\underline{B}}$. From the definitions, one has

$$P_{J,\underline{X}}[K\times C] = P_{J\otimes N}\{[K\times\mathbb{R}^I] \cap [\underline{X}_{\Phi} \in C]\} = \int_K P_J(du)P_N\{v: \underline{X}_{\Phi}[{}^u_v] \in C\}.$$

This shows, in particular, that $P_N\circ[\underline{X}^u_{\Phi}]^{-1}$ is a version of the conditional law of $\underline{X}$ given $J = u$, where $\underline{X}^u_{\Phi}(v) = \underline{X}_{\Phi}[{}^u_v]$.

Now, since

$$\underline{X}_{\Phi}(\underline{f},t) = \int_0^t \Gamma(dx)\underline{\tau}_{\Phi}\left[\begin{bmatrix} q_1 \\ \underline{X}_{\Phi} \end{bmatrix}(\underline{f}),x\right] + \underline{B}_{\Phi}(\underline{f},t),$$

by fixing the first component, u, in the argument of $\underline{X}_{\Phi}$, one has a stochastic differential equation whose solution, $\underline{X}^u_{\Phi}$, has law $P_N\circ[\underline{X}^u_{\Phi}]^{-1}$ absolutely continuous with respect to $P_N\circ\underline{B}_{\Phi}^{-1}$, which means that $P_{\underline{X}|J=u}$ is absolutely continuous with respect to $P_{\underline{B}}$. Thus

$$P_{J,\underline{X}}(K\times C] = \int_K P_J(du)\int_C P_{\underline{B}}(d\underline{c})\frac{dP_{\underline{X}|J=u}}{dP_{\underline{B}}}(\underline{c}) = \int_{K\times C} d[P_J\otimes P_{\underline{B}}]([{}^u_{\underline{c}}])\frac{dP_{\underline{X}|J=u}}{dP_{\underline{B}}}(\underline{c})$$

$$= \int_{K\times C} dP_{J,\underline{B}}([{}^u_{\underline{c}}])\frac{dP_{\underline{X}|J=u}}{dP_{\underline{B}}}(\underline{c}).$$

Moreover,

$$\frac{dP_{\underline{X}|J=u}}{dP_{\underline{B}}}(\underline{c}) = \exp\left\{\left[\frac{1}{\gamma^2(\underline{c},1)}\right]\left[\int_0^1 \langle \underline{\tau}_\Phi([{}^u_{\underline{c}}],x),\ \underline{\Pi}(\underline{c},dx)\rangle_{\mathbb{R}^M} - \tfrac{1}{2}\int_0^1 \langle \underline{\tau}_\Phi([{}^u_{\underline{c}}],x),\ \Gamma(dx)\underline{\tau}_\Phi([{}^u_{\underline{c}}],x)\rangle_{\mathbb{R}^M}\right]\right\}$$

Note that $h: (u,\underline{c}) \longrightarrow \frac{dP_{\underline{X}|J=u}}{dP_{\underline{B}}}(\underline{c})$ is product-measurable [22].

(d) Computation of $D_{\underline{X},\underline{B}}$. One begins again with the representation

$$\underline{X}(\omega,t) = \int_0^t \Gamma(dx)\underline{\tau}(\underline{x}[\omega],x) + \underline{B}(\omega,t).$$

Let $\bar{\underline{\tau}}$ have components $\tau_i \frac{d\Gamma_i}{d\Gamma_1}$, $1 \le i \le M$. Assumption 4.1.c then yields

$$E\int_0^1 \Gamma_1(dx)\|\bar{\underline{\tau}}(\underline{x}[\omega],x)\|^2_{\mathbb{R}^M} < \infty.$$

Using Proposition 3 and Memin [19], one obtains that $(\underline{X}_t)$ has a representation of the form

$$\underline{X}(\omega,t) = \int_0^t \Gamma(dx)\underline{\tau}_C(\underline{X}[\omega],x) + \underline{B}^X(\omega,t)$$

where: $\underline{\tau}_C$ is predictable for $\mathscr{C}^M$;

$E_{P_{\underline{X}}}\int_0^1 \langle \underline{\tau}_C(\underline{c},x),\ \Gamma(dx)\ \underline{\tau}_C(\underline{c},x)\rangle_{\mathbb{R}^M} < \infty$;

$(\underline{B}^X_t)$ is a spherically invariant martingale for $\sigma(\underline{X})$ such that

$\langle \underline{B}^X \rangle = A^2\Gamma$.

Consequently, with respect to $P_{\underline{X}}$,

$$D_{\underline{X},\underline{B}}(\underline{c}) = \frac{dP_{\underline{X}}}{dP_{\underline{B}}}(\underline{c}) = \exp\left[\frac{1}{\gamma^2(\underline{c},1)}\left\{\int_0^1 \langle \mathcal{I}_C(\underline{c},x),\ \underline{\Pi}(\underline{c},dx)\rangle_{\mathbb{R}^M} - \tfrac{1}{2}\int_0^1 \langle \mathcal{I}_C(\underline{c},x),\ \Gamma(dx)\mathcal{I}_C(\underline{c},x)\rangle_{\mathbb{R}^M}\right\}\right],$$

where $\mathcal{I}_C(\underline{X}[\omega],t)$ is a version of the conditional expectation $E[\mathcal{I}(\underline{x}[\cdot],t)|\sigma_t(\underline{X})]$.

## 4.4. Expression for $I(J,\underline{X})$

One has by the change of variable formula,

$$I(J,\underline{X}) = \int_{\mathbb{R}^{[0,1]}\times C_0^M[0,1]} \ln\left\{\frac{dP_{J,\underline{X}}}{dP_J\otimes P_{\underline{X}}}([{}^f_{\underline{c}}])\right\}dP_{J,\underline{X}}([{}^f_{\underline{c}}]) = E\ \ln\left\{\frac{dP_{J,\underline{X}}}{dP_J\otimes P_{\underline{X}}}\begin{bmatrix}J(\cdot)\\ \underline{X}(\cdot)\end{bmatrix}\right\}$$

$$= E\ \ln\left\{D_{J,\underline{X},\underline{B}}\begin{bmatrix}J(\cdot)\\ \underline{X}(\cdot)\end{bmatrix}\right\} - E\ \ln\left\{D_{\underline{X},\underline{B}}(\underline{X}(\cdot))\right\}.$$

Now
$$\ln D_{J,\underline{X},\underline{B}}\left(\begin{bmatrix}J(\omega)\\ \underline{X}(\omega)\end{bmatrix}\right) = \frac{1}{\gamma^2(\underline{X}[\omega],1)}\int_0^1 \left\langle \mathcal{I}_\Phi\left[\begin{bmatrix}J(\omega)\\ \underline{X}(\omega)\end{bmatrix},x\right],\ \underline{X}(\omega,dx)\right\rangle_{\mathbb{R}^M} - \tfrac{1}{2}\int_0^1 \left\langle \mathcal{I}_\Phi\left[\begin{bmatrix}J(\omega)\\ \underline{X}(\omega)\end{bmatrix},x\right],\ \Gamma(dx)\mathcal{I}_\Phi\left[\begin{bmatrix}J(\omega)\\ \underline{X}(\omega)\end{bmatrix},x\right]\right\rangle_{\mathbb{R}^M}.$$

Now $\gamma^2(\underline{X}[\omega,1]) = \gamma^2(\underline{B}[\omega],1) = A^2(\omega)$, and $\mathcal{I}_\Phi\left[\begin{bmatrix}J(\omega)\\ \underline{X}(\omega)\end{bmatrix},t\right] = \underline{s}(\omega,t)$, so that

$$\ln D_{J,\underline{X},\underline{B}}\left(\begin{bmatrix}J(\omega)\\ \underline{X}(\omega)\end{bmatrix}\right) = \frac{1}{A^2}\left\{\int_0^1 \langle \underline{s}(\omega,x),\underline{X}(\omega,dx)\rangle_{\mathbb{R}^M} - \tfrac{1}{2}\int_0^1 \langle \underline{s}(\omega,x),\Gamma(dx)\underline{s}(\omega,x)\rangle_{\mathbb{R}^M}\right\}$$

$$= \frac{1}{2A^2}\|S\|_N^2 + \frac{1}{A^2}\int_0^1 \langle \underline{s}(\omega,x),\underline{B}(\omega,dx)\rangle_{\mathbb{R}^M}.$$

Similarly,

$$\ln D_{\underline{X},\underline{B}}(\underline{X}(\omega)) = \frac{1}{2A^2}\|\hat{S}\|_N^2 + \frac{1}{A^2}\int_0^1 \langle \hat{\underline{s}}(\omega,x),\underline{B}(\omega,dx)\rangle_{\mathbb{R}^M}.$$

where $\hat{S}$ and $\hat{\underline{s}}$ denote appropriate versions of the conditional expectations of S and $\underline{s}$ respectively, the filtration being $\underline{\sigma}(\underline{X})$.

So, under the assumption that

$$E \frac{1}{A^2}\|S\|_N^2 < \infty,$$

one has

$$I(J,Y) = \tfrac{1}{2}\left\{E \frac{1}{A^2}[\|S\|_N^2 - \|\hat{S}\|_N^2]\right\}.$$

From this expression, an upper bound for the mutual information of the feedback channel is $\tfrac{1}{2}\, E \frac{1}{A^2}\|S\|_N^2$. If the signal is contrained by $E\|S\|_N^2 \le P_0$, then an upper bound on the capacity of the no-feedback channel (S and A independent) is obviously $\tfrac{1}{2}\, P_0 E(\frac{1}{A^2})$. This is also a lower bound [3], so that the no-feedback capacity (subject to $E\|S\|_N^2 \le P_0$) is $\tfrac{1}{2}\, P_0 E(\frac{1}{A^2})$. The lower bound can be shown in several ways. For example, one can note that it is sufficient to consider $I(J,X)$, where

$$X(\omega,t) = \int_0^t J(\omega,s)ds + A(\omega)W(\omega,t),$$

$E \int_0^1 J^2(\omega,s)ds \le P_0$, and $(W_t)$ is the standard Wiener process. $\|X\|^2 \equiv \int_0^1 X^2(t)dt$. Choosing a sequence $(J_t^n)$ of zero-mean Gaussian processes independent of A and $(W_t)$, with $(J_t^n)$ having covariance function $P_0 e^{-n|t-s|}$, one can show that $E\|J^n\|^2 = P_0$ while $\lim_n E\|\hat{J}^n\|^2 = 0$. That is, from [15], $E[\|\hat{J}^n\|^2|A] \le n^{-1}P_0$ a.s., so that $E\|\hat{J}^n\|^2 \to 0$.

# References

1. C.R. Baker, Absolute continuity and applications to information theory, *Probability in Banach Spaces*, Oberwolfach 1975, A. Beck, Ed., Lecture Notes in Mathematics No. 526, pp.1-11. Springer-Verlag: Berlin (1976).

2. __________, Capacity of the mismatched Gaussian channel, *IEEE Trans. Inf. Theory*, *IT-33*, 802-812 (1987).

3. __________ and A.F. Gualtierotti, Signal detection and channel capacity for spherically-invariant processes, *Proc. 23rd IEEE Conf. on Decision and Control*, 1444-1446 (1984).

4. ____________________________________, Discrimination with respect to a Gaussian process, *Prob. Th. and Related Fields*, *71*, 159-182 (1986).

5. ____________________________________, Likelihood ratios and signal detection for non-Gaussian processes, Chapter 6, *Stochastic Processes in Underwater Acoustics*, C.R. Baker, Ed., Lecture Notes in Control and Information Sciences No. 85, pp. 154-180. Springer-Verlag: Berlin (1986).

6. M. Bouvet, Détection en environnement non-gaussien: différentes approches et utilisation des modèles de mixture, *Traitement du Signal 4*, 101-113 (1987).

7. J. Bretagnolle, D. Dacunha-Castelle and J.L. Krivine, Lois stables et espaces $L^p$, *Ann. Inst. Henri Poincare, Ser. B2*, 231-259 (1966).

8. H. Cramér, On some classes of non-stationary stochastic processes, *Proc. Fourth Berkeley Symp. on Math. Stat. and Prob.*, 2, 57-77 (1961).

9. C. Dellacherie and P.A. Meyer, *Probabilities et Potentiel*. Hermann: Paris (1980).

10. J.L. Doob, *Stochastic Processes*. Wiley: New York (1953).

11. T. Hida, Canonical representations of Gaussian processes and their applications, *Mem. Coll. Science, Univ. Kyoto* 33A, 109-155 (1960).

12. M. Hitsuda and S. Ihara, Gaussian channels and the optimal coding, *J. Multivar. Anal.* 9, 106-118 (1975).

13. S.T. Huang and S. Cambanis, Spherically invariant processes: their non linear structure, discrimination and estimation, *J. Multivar. Anal.* 9, 59-83 (1979).

14. K. Ito, The topological support of Gaussian measures on Hilbert space, *Nagoya Math. J.* 38, 181-183 (1970).

15. T.T. Kadota, M. Zakai and J. Ziv, Mutual information of Gaussian channels with and without feedback, *IEEE Trans. Information Theory 17*, 368-371 (1971).

16. G. Kallianpur, *Stochastic Filtering Theory*. Springer-Verlag: New York (1980).

17. E. Lenglart, Tribus de Meyer et théorie des processus, *Seminaire des Prob. XIV*, 1978/79, Lectures Notes in Mathematics No. 784, J. Azéma and M. Yor, Eds. Springer-Verlag: Berlin (1980).

18. R.S. Lipster and A.N. Shiryayev, *Statistics of Random Processes*. Springer-Verlag: Berlin (1977).

19. J. Mémin, Sur quelques problèmes fondamentaux de la théorie du filtrage, Thèse (troisième cycle), *U.E.R. Mathematiques et Informatique*. University of Rennes (1974).

20. M. Métivier, *Semimartingales*. De Gruyter: Berlin (1982).

21. B. Picinbono and G. Vezzosi, Détection d'un signal certain dans un bruit non stationnaire et non gaussien, *Annales des Telecommunications* 25, 433-439 (1970).

22. C. Stricker and M. Yor, Calcul stochastique dépendant d'un paramètre, *Z. Wahrscheinlichkeitstheorie verw. Gebiete* 45, 109-133 (1978).